

Interactive transmission processing for large images in a resource-constraint mobile wireless network

Yi Zhuang¹ · Nan Jiang² · Hua Hu³ ·
Dickson K. W. Chiu⁴ · Qing Li⁵

Received: 2 February 2016 / Revised: 11 July 2016 / Accepted: 14 September 2016 /
Published online: 21 November 2016
© Springer Science+Business Media New York 2016

Abstract In the state-of-the-art methods for (large) image transmission, no user interaction behaviors (e.g., user tapping) can be actively involved to affect the transmission performance (e.g., *higher image transmission efficiency with relatively poor image quality*). So, to effectively and efficiently reduce the large image transmission costs in resource-constraint mobile wireless networks (*MWN*), we design a content-based and bandwidth-aware Interactive large Image Transmission method in *MWN*, called the *IIT*. To the best of our knowledge, this is the first study on the interactive image transmission. The whole transmission processing of the *IIT* works as follows: before transmission, a preprocessing step computes the optimal and initial image block (*IB*) replicas based on the image content and the current network bandwidth at the sender node. During transmission, in case of unsatisfied transmission efficiency, the user's anxiety to preview the image can be implicitly indicated by the frequency of tapping the screen. In response, the transmission resolutions of the candidate *IB* replicas can be dynamically adjusted based on the user anxiety degree (*UAD*). Finally, the candidate *IB* replicas are transmitted with different priorities to the receiver for reconstruction and display. The experimental results show that the performance of our approach is both efficient and effective, minimizing the response time by decreasing the network transmission cost while improving user experiences.

Keywords Interactive transmission · Multi-resolution · Mobile network · User anxiety degree

✉ Yi Zhuang
zhuang@mail.zjgsu.edu.cn

¹ College of Computer & Information Engineering, Zhejiang Gongshang University, Hangzhou, People's Republic of China

² Hangzhou First People's Hospital, Hangzhou, People's Republic of China

³ School of Computer, Hangzhou Dianzi University, Hangzhou, People's Republic of China

⁴ Faculty of Education, The University of Hong Kong, HKSAR, Hong Kong, People's Republic of China

⁵ Department of Computer Science, City University of Hong Kong, HKSAR, Hong Kong, People's Republic of China

1 Introduction

In the state-of-the-art methods for (large) image transmission, in most cases, a user has to passively wait till completed transmission to see the entire image, and no user interaction behaviors (e.g., user tapping, etc.) are considered to affect the transmission performance (e.g., higher image transmission efficiency with relatively poor image quality). Image transmission in a resource-constraint mobile wireless network (*MWN*) usually have three main characteristics: 1) **Mobility of users**: most users using the *MWN* are constantly moving which means the spatial position of each user varies with time; 2) **Lower processing power**: the processing and battery capacities of most mobile client devices (e.g., cell phone, PDA, etc) are very limited, which motivates us to devise an energy-efficient technique to support the image transmission with lower costs; 3) **Instability and heterogeneity of the *MWN***: the bandwidths in the *MWN* are often unstable, that means, some nodes may be down or connected intermittently to the network. The bandwidth between any two nodes in the *MWN* may be different according to the variance of time.

For the transmission and browsing of large images in *MWN*, the network transmission cost of such images will consume a large percentage of the overall interaction time [19]. So, the reduction of the transmission cost, especially in *MWN*, is critical to the transmission performance improvement. Since an image usually contains some salient objects with corresponding regions called *Region of Salient Objects* (RSO), the total image data size can be effectively reduced if the pixel resolutions of the non-RSO region can be moderately reduced based on the network bandwidth. This will not affect the user experience in watching, but the total transmission cost can often be decreased significantly (i.e., higher image transmission efficiency with inferior image quality in unimportant regions). Moreover, in the state-of-the-art methods, no human interactions (e.g., *tapping the receiver screen*) are considered during large image transmission processing, which may cause a user to wait for a long time if the current network bandwidth is limited and unstable, especially in *MWN* environment. So the user experiences are bad in this case since he is anxious to watch the image. If the pixel resolutions of non-RSO replicas can be interactively adjusted by the receiver users during the transmission processing, not only the transmission performance is improved, but also the user experiences can be significantly enhanced. Based on the above motivations, we propose a content-based and bandwidth-aware Interactive Image Transmission (*IIT*) scheme that not only considers the image content and the network bandwidth, but also user interactions. For an active image transmission processing, exploring user interactions to optimize the transmission processing and improve the user experiences is a new research topic, which has received little attention so far.

The basic idea behind the transmission scheme is that given a transmission image I_S , its corresponding RSOs are first automatically detected by the approach of Girshick et al. [11]. As the RSOs are main contents of the image, they are critically important to the *IIT* processing in which their pixel resolutions are kept original. However, for the rest of the image (i.e., non-RSO), both its pixel resolution and transmission priority are lower than that of the RSOs so that the key part of the image can be transmitted and displayed first. Note that, the optimal resolution of the non-RSOs can be derived based on factors such as *current network bandwidth*, *corresponding size*, *user interactions*, etc. Then, the *IIT* processing of I_S becomes an active transmission of the image blocks (IB) in I_S that has been partitioned in the preprocessing step (ref. Section 3.2). The IB replicas with different resolutions and transmission priorities are stored at the slave node level (N_L). During the transmission processing, in case of unsatisfied transmission efficiency, the user's anxiety to watch the image can be implicitly indicated by the frequency of tapping the receiver's screen. Then, the IB replicas with lower pixel resolutions

are chosen as candidate ones for transmission, reconstruction, and display at the receiver node. For those IBs with lower resolutions, the user can also request new IB replicas with higher resolutions by tapping the corresponding region on the screen.

The challenges of designing such an innovative high performance interactive image transmission method include the five main aspects:

- 1) **Applying user interaction (i.e., screen tapping) to image transmission:** as user interaction (i.e., tapping the screen) means the user hopes the image transmission can be completed earlier, the problem is how to establish a relationship between the frequency of screen tapping and the resolution of non-RSO part.
- 2) **High computation cost in image transmission:** most images are characterized by high pixel resolution, high dimension, and large scale. So the transmission cost of such images is very high.
- 3) **Mobility of MWN users:** as most users in the *MWN* are often moving which means the spatial position of each user varies with time, how to perform an optimal data placement is also a challenging issue.
- 4) **Resource-Constraint MWN:** the power capacities of the mobile devices are very limited. The display resolutions of such mobile devices are often low. Furthermore, the bandwidth in the *MWN* is limited, how to transmit such a large image in the resource-constraint *MWN* is challenging.
- 5) **Instability and heterogeneity of the MWN:** the nodes in the *MWN* are often instable, that means, some nodes may be down or connected intermittently to the network. The bandwidth between any two nodes in the *MWN* varies with time. There is no guarantee that the total response time of each transmission will be similar.

To address the above challenges, we propose two enabling techniques in the *IIT*, i.e., a *multi-resolution-based interactive RIB replica selection scheme* and an *optimal IB replica placement scheme*. We have implemented the *IIT* method and extensive experiments indicate that our approach is specifically suitable for large image transmission in a relatively low network bandwidth with much enhanced user interaction experiences. Our contributions can be summarized as follows:

- We introduce a framework of an interactive image transmission scheme in the resource-constraint mobile network (*IIT*).
- We propose an interactive user model to support the *IIT* processing.
- We present a multi-resolution-based interactive RIB replica selection scheme to adaptively reduce the communication cost in the *MWN* environment and improve user experiences.
- We design an optimal IB replica placement scheme to adaptively reduce the storage cost.

The remainder of the paper is organized as follows. Section 2 reviews related techniques and Section 3 presents preliminary definitions and preprocessing step. After that, to effectively and efficiently facilitate the interactive image transmission processing, we present two enabling techniques, i.e., a *multi-resolution-based interactive RIB replica selection scheme* and an *optimal IB replicas placement scheme* in Sections 4 and 5, respectively. In Section 6, we propose an interactive image transmission scheme called the *IIT*. In Section 7, we perform comprehensive experiments to evaluate the efficiency of our proposed approach before we conclude the paper in Section 8.

2 Related work

In this section, we review key related works on image data transmission techniques. Image data transmission techniques have been studied for about 20 years [1, 8, 16, 18, 19, 21]. The state-of-the-art methods can be mainly divided into two categories: 1) *improvement design of transmission protocol* [8, 9, 16, 19, 21–23]; and 2) *image data encoding and compression* [3–7, 9, 17, 18, 22, 26].

Charles et al. [8] first proposed a wireless image data transmission method from end to end, and provided experimental analysis. As images are usually transmitted across the Internet using a lossless protocol such as TCP/IP, lossless protocols require retransmission of lost packets, which substantially increases transmission time. John et al. [16] presented a fast lossy Internet image transmission scheme (FLIIT) for compressed images which eliminates retransmission delays by shielding important portions of the image with redundancy bits. They described a joint source and channel coding algorithm for images which minimizes the expected distortion of transmitted images. After that, Raman et al. [21] proposed an image transmission protocol called ITP. Comparing with the traditional TCP protocol, the ITP is more suitable for image data transmission. Gao et al. [9] presented a robust image transmission scheme for wireless channels based on compressive sensing. Due to the high packet error rates and the need for retransmission, Aziz et al. [23] has designed a novel architecture for energy efficient image processing and communication over wireless sensor networks. Recently, Maani et al. [19] introduced a parallelism to provide an efficient method of medical image transmission based on parallel TCP connection.

Since most image transmission methods use the same pixel interpolation scheme for the entire picture, without considering the differences in different parts, Chang et al. [7] presented a progressive image transmission (PIT) scheme which transmits the most significant part of a picture, followed by less important parts. Lin et al. [18] presented a compound image compression algorithm for real-time applications of computer screen image transmission called *Shape Primitive Extraction and Coding* (SPEC). Ruiz et al. [22] designed an image compression algorithm to support progressive image transmission. Available Pit mechanisms and systems can be categorized into spatial domain [5], and pyramid-structured progressive transmission [17]. In the transform domain, an image undergoes block compression and the transformed coefficients are transmitted progressively in a relative importance order (e.g., Progressive JPEG). Alternatively, a germinal and instinctive method for progressive image transmission in the spatial domain is the *Bit Plane Method* (BPM) [4, 26]. In this method, the final transmitted image is the same as the original. However, its high transmission bit rate is a major disadvantage of BPM.

Due to the drawback of BPM, lossy PIT techniques have received more attention. To provide a fast PIT scheme, Chang et al. [6] improved the BPM method by color guessing called the guessing by neighbors (GBN), which uses interleaved pixels for transmission. Fifty percent of the pixels are transmitted while the other 50 % were “guessed”. Sun et al. [24] proposed a progressive image transmission system over wireless channels by combining *joint source-channel coding* (JSCC), space-time coding, and *orthogonal frequency division multiplexing* (OFDM). Based on the Reed-Solomon coding scheme, Boluk et al. [3] proposed a robust image transmission over wireless sensor networks. Victor et al. [27] devised a 3-D scalable image compression method with optimized volume of interest coding. Arslan et al. [2] has proposed a generalized unequal error protection LT codes for progressive data transmission. Although Hu et al. [12] introduced an attention model based progressive image transmission method in which the *region of interest* (ROI) in an image can be transmitted with high-resolution in priority, the human interaction can not be allowed to affect the transmission

efficiency. Wu and Abouzeid [28] devised a power aware image transmission in energy constrained wireless networks. Gelogo and Kim [10] reviewed the compressed images transmission issues and solutions. Zhuang et al. [30] have proposed a content-aware and multi-resolution-based image transmission scheme in which only two factors, i.e., the image content and the network bandwidth are considered to optimize the transmission processing. Recently, Xua et al. [29] proposed an adaptive FEC coding and cooperative relayed wireless image transmission.

Different from the above state-of-the-art methods, the paper proposes an interactive image data transmission method in which the three factors, i.e., *the receiver's interaction behaviors*, *the image content*, and *current network bandwidth* are holistically analyzed to obtain an optimal transmission pixel resolution. To the best of our knowledge, this is the first attempt to improve the transmission performance from the perspectives of the user interactions and the image contents.

3 Preliminaries and preprocessing

3.1 Preliminaries

The list of symbols to be used throughout the rest of this paper is summarized in Table 1.

Definition 1. A mobile wireless network (MWN) is a graph which is represented by a triplet:

$$MWN := \langle N, E, T \rangle \quad (1)$$

where N refers to the set of nodes, E refers to a set of edges representing the network bandwidths for data transmission at time T .

In Fig. 1, due to the instability and heterogeneity of the MWN environment, the bandwidth of any two nodes in MWN may be different and vary with time. In addition, the data transmission distance in the mobile network is limited.

Table 1 Meaning of symbols used

Symbols	Meaning
Ω	A set of images
I_S	A transmission image and $I_S \in \Omega$
RSO_i	The i -th RSO in an image and $i \in [1, RSO]$
R	The rest part of the image I_S (i.e., non-RSO)
IF	Image fragment (i.e., RSO and R)
IB	Image block
D_L, D_U	The lower and upper bound of the dots per inch
E_L, E_U	The lower bound and upper bound of the network bandwidth
$ RSO $	The number of RSOs in an image
U_R	A receiver user
$\lceil * \rceil$	The integral part of *
δ	Granularity value for the time threshold
Δ	Granularity value both for network bandwidth and the image pixel resolution

Definition 2. The nodes in the MWN can be logically divided into three categories: the sender node (N_S), the slave node (N_L), and the receiver node (N_R), formally denoted as $N = N_S \cup N_L \cup N_R$, where

- N_S is responsible for obtaining an optimal IB transmission pixel resolution based on the collection and analysis of the network bandwidth and user interaction behaviors;
- N_L is responsible for: 1) storing the IB replicas with different pixel resolution and transmission priorities, and 2) sending the IB replicas to the receiver;
- N_R is responsible for: 1) receiving, reconstructing, and displaying the images; 2) sending user interactions (e.g., tapping the screen) to N_S .

For each image, in most cases, there exist some *salient objects* that users are interested in. The regions of such salient objects are called *region of salient objects (RSO)*.

Definition 3. A region of salient object (RSO) in an image can be modeled by a four-tuple:

$$RSO_i := \langle i, S, pos, dpi \rangle \tag{2}$$

where i is the ID number of the RSO, S is the area value of the RSO, pos is the coordinate position of the RSO in the image, and dpi refers to the dots per inch (dpi) for the RSO.

Definition 4. A non-RSO part of an image, denoted as R , can be modeled by a two-tuple:

$$R := \langle S, dpi \rangle \tag{3}$$

where S is the area value of the R , dpi refers to the dots per inch for the R .

Based on Definitions 3 and 4, in Fig. 2, there exists one RSO (e.g., RSO_1) and one non-RSO (e.g., R) in the image. The corresponding RSO area can be preliminarily detected by the approach proposed by Girshick et al. [11] and drew by a black rectangle line.

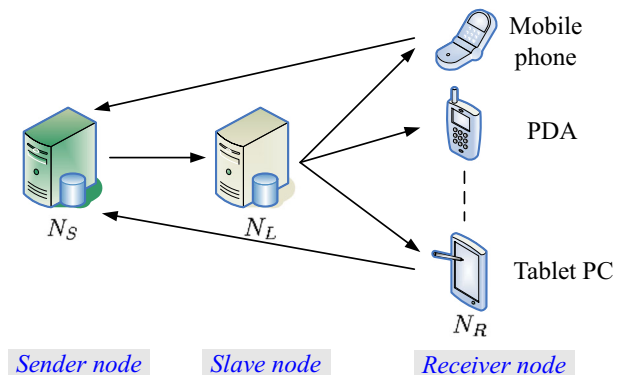
As mentioned before, for an image, it can be equally partitioned into blocks called *image blocks*.

Definition 5 (IMAGE BLOCK). Given an image block IB_i , it can be modeled by a five-tuple:

$$IB_i := \langle i, S, pos, TP, dpi \rangle \tag{4}$$

where i refers to the ID number of the IB , S is the area value of the IB , pos is the coordinate position of the IB in the image, TP refers to the transmission priority of the IB ($TP \in [0, 1]$), and dpi means the corresponding pixel resolution of the IB ;

Fig. 1 Three-layer architecture in a MWN



Note that, there are two kinds of IBs (i.e., *SIB* and *RIB*) in terms of their positions in the image, which are defined below:

Definition 6 (SIB). A *SIB* is an image block which intersects with a *RSO* or is contained by a *RSO*, formally represented as:

$$SIB = \{IB_i | IB_i \cap RSO_j \neq \emptyset\} \quad (5)$$

where $i \in [1, \alpha + \beta]$ and $j \in [1, |RSO|]$.

Definition 7 (RIB). A *RIB* is an image block contained by the non-*RSO* part(*R*) of an image, formally denoted by:

$$RIB = \{IB_i | IB_i \cap R = IB_i\} \quad (6)$$

where $i \in [1, \alpha + \beta]$.

Based on Definitions 6 and 7, also in Fig. 2, the *RSO* has 16(4×4) *SIBs* (i.e., IB_{23} , IB_{24} , IB_{25} , IB_{26} , IB_{33} , IB_{34} , IB_{35} , IB_{36} , IB_{43} , IB_{44} , IB_{45} , IB_{46} , IB_{53} , IB_{54} , IB_{55} and IB_{56}). The rest of the *IBs* are *RIBs*.

Our proposed *IIT* method aims at transmitting a large image adaptively and efficiently in a limited network bandwidth and a user-preferred time (T_θ) that is dynamically set through the user interactions (e.g., *tapping the receiver screen*) during the transmission processing.

3.2 Preprocessing

As mentioned above, in the preprocessing step of the *IIT*, for each image I_i in Ω , it can be physically and equally partitioned into several *IBs*. The corresponding *SIBs* and *RIBs* are placed at the slave node level with different pixel resolutions and transmission priorities. Algorithm 1 details the initial *IB* replicas placement processing.

Algorithm 1. Initial *IB* replicas placement()

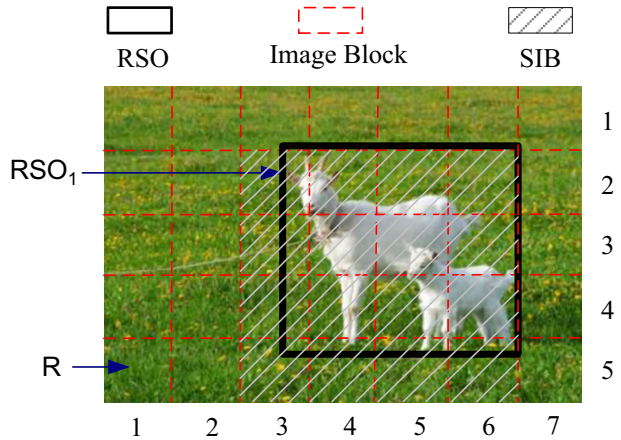
Input: n images in Ω

Output: the initial *IB* replicas placement

1. **for** each image I_i in Ω **do**
 2. its corresponding *RSOs* are identified manually or automatically;
 3. the image is equally partitioned into $(\alpha + \beta_i)$ *IBs*;
 4. the *RIB* replicas with different resolutions are stored at the slave node based on A_{mi} ;
 5. the *SIBs* are stored at the slave node with a fixed resolution;
 6. **end for**
-

Generally speaking, in most of the state-of-the-art image data transmission schemes, an image is transferred as a whole object in which the transmission priorities of all *IBs* are equal. Thus, the important *RSOs* in the image are often displayed later than the rest of the image. For images with high pixel resolutions, such methods further leads to the increase of transmission failure rate. Once transmission failure occurs, the image needs to be re-transmitted, resulting in an even higher transmission overhead. To overcome this technical bottleneck, we propose a *Transmission Priority (TP) Assignment for IB replicas scheme* called *TPA* to support the successive and robust transmission of the large image data. The *TP* of each *IB* is defined in Eq. (7).

Fig. 2 One RSO in an image



$$IB_i.TP = \begin{cases} 1, & \text{if } IB_i \cap RSO_j = IB_i \\ 0.5, & \text{if } IB_i \cap RSO_j \neq \phi \\ 0, & \text{Otherwise} \end{cases} \tag{7}$$

According to the different *TPs* of the *IBs*, the *IBs* can be transmitted in terms of their corresponding *TPs* in a descending order, which not only ensure the robustness of data transmission but guarantee that the important information can be transmitted in advance.

4 Multi-resolution-based interactive RIB replica selection scheme

As mentioned before, high pixel resolutions of digital images usually lead to a large data size accordingly. It is non-trivial to transmit an image of such a big size to the receiver nodes directly, especially in a resource-constraint mobile network environment in which the network bandwidth is limited and unstable.

Based on the above analysis, in this section, we propose a *Multi-resolution-based Interactive image data Replica selection scheme* (MIR) by uniformly analyzing the relationship of the *image content, network bandwidth, and human interactions*.

4.1 Modeling user interaction behaviors

To facilitate the interactive image transmission processing, it is critical to modeling the user’s interactive behaviors. Specifically, for a receiver user U_R , his anxiety can be modeled by the times of tapping the screen during a certain short time interval (ΔT). So in this subsection, we first give a definition of the *user anxiety degree* (UAD).

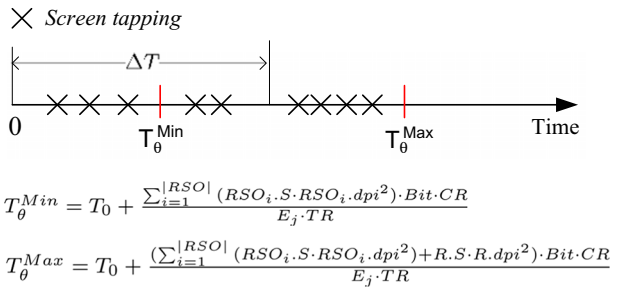
Definition 8 (USER TAPPING). A user tapping (*UT*) can be modeled by a triple-tuple:

$$UT_i := \langle i, UID, T \rangle \tag{8}$$

where *UID* means the user ID and *T* refers to the time when the tapping is performed.

Definition 9 (USER ANXIETY DEGREE). For a user U_i , his UAD can be defined as the times for tapping on the screen during a short time interval (i.e., ΔT), formally represented as:

Fig. 3 Selection of ΔT



$$UAD_i = \sum_{UT_i, T \in \Delta T} 1 \tag{9}$$

where UT_i, T means the time when the i -th user tapping is performed and $\Delta T \leq T_{\theta}^{Max}$.¹

Based on Definition 9, the larger the UAD is, more anxious the user is. Next, we establish the relationship between a transmission deadline (T_{θ}) and the UAD . Suppose that the image transmission processing can be finished in T_{θ} that is dynamically adjusted based on the users tapping the screen, so for T_{θ} , it can be represented below:

$$T_{\theta} = T_{\theta}^{Min} + \frac{T_{\theta}^{Max} - T_{\theta}^{Min}}{\delta} \cdot k \tag{10}$$

where $T_{\theta} \in [T_{\theta}^{Min}, T_{\theta}^{Max}]$ (see Fig. 3), T_{θ}^{Min} is a minimal time threshold (i.e., $T_{\theta}^{Min} = T_0 + \frac{\sum_{i=1}^{|RSO^1|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2) \cdot Bit \cdot CR}{E_j \cdot TR}$), T_{θ}^{Max} is a maximal one (i.e., $T_{\theta}^{Max} = T_0 + \frac{(\sum_{i=1}^{|RSO^1|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2) + R \cdot S \cdot R \cdot dpi^2) \cdot Bit \cdot CR}{E_j \cdot TR}$). δ is a granularity value that can be tuned by users and $k \in [1, \delta]$.

Based on Eq.(10), more anxious the user is, the smaller T_{θ} is. Since the user anxiety can be measured by the UAD defined in Definition 9. Moreover, k is proportional to T_{θ} in Eq.(10). Therefore, the larger the user’s UAD is, the smaller k is. Based on this observation, the relationship between k and UAD can be approximately represented in Fig. 4.

With the increase of the UAD , k is decreasing gradually. k and UAD can be approximately modeled by Eq.(11):

$$k = \left\lceil \frac{\delta}{UAD + 1} \right\rceil \tag{11}$$

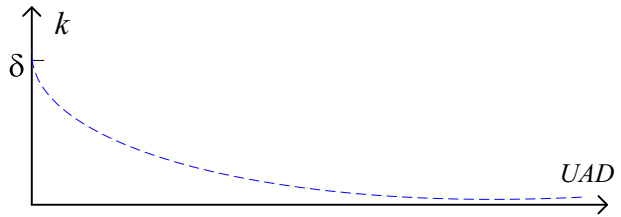
Based on Eqs. (10–11), T_{θ} can be derived in Eq. (12):

$$T_{\theta} = T_{\theta}^{Min} + \frac{T_{\theta}^{Max} - T_{\theta}^{Min}}{\delta} \cdot \left\lceil \frac{\delta}{UAD + 1} \right\rceil$$

$$\approx T_{\theta}^{Min} + \frac{T_{\theta}^{Max} - T_{\theta}^{Min}}{UAD + 1} \tag{12}$$

¹ T_{θ}^{Max} is a maximal transmission time which is defined in Eq.(10).

Fig. 4 k and UAD



4.2 Optimal ID of RIB replica

The basic idea of the MIR method is that for a certain image, its transmission pixel resolution needs to be adjusted based on the variance of the network bandwidth. Specifically, with higher network bandwidth, a higher-resolution image can be transferred in a reasonable short period of time (T_θ) (ref. Eq.(12)), where T_θ can be adjusted by a user interactively during the transmission processing. On the contrary, in order to get a shorter response time, a lower-resolution version of the same image can be sent to the receiver node with lower network bandwidth.

Although reducing the pixel resolution of the whole image can reduce the transmission overhead, some salient objects (i.e., RSO), however, cannot be clearly viewed. Therefore, we adjust the resolutions of the non-RSO part in the image just moderately based on the network bandwidth and the user interactions. Compared with the non-RSO part(R) of the image, the RSOs is best displayed with the original pixel resolution. For the non-RSO area, as shown in Fig. 5a, it can be equally partitioned into some *RIBs* by yellow dash lines.

Based on the above analysis, the objective of our method is to get a trade-off between the quality of image and the transmission cost under different resolutions and available network bandwidth.

Suppose that the image transmission processing can be finished in T_θ set by user, so we have:

$$T_T = T_0 + \frac{Size(I_i)}{BWidth(E_j)} \leq T_\theta \tag{13}$$

where

- $Size(I_S)$ is the data size of I_S , represented as: $Size(I_S) = (\sum_{i=1}^{RSO} (RSO_i \cdot S \cdot RSO_i \cdot dp_i^2) + R \cdot S \cdot R \cdot dp^2) \cdot Bit \cdot CR$, where Bit means color bit, and Bit can be 8, 16, or 24, CR is an image compression ratio and $CR \in [0,1]$;
- T_0 is the start-up transmission time;

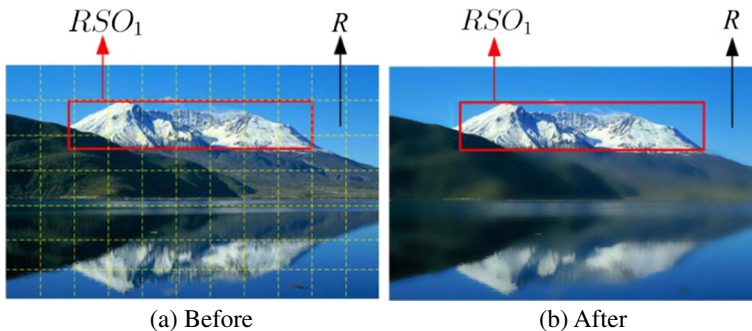


Fig. 5 Comparison of the pixel resolutions of a same image

- $BWidth(E_j)$ is a real network bandwidth, denoted as $BWidth(E_j) = E_j \cdot TR$, where E_j is a theoretical network bandwidth, TR means a attenuation rate of the bandwidth, and $TR \in [0, 1]$;
- T_θ is a time threshold and $T_\theta \in [T_\theta^{Min}, T_\theta^{Max}]$, where the definitions of T_θ^{Min} and T_θ^{Max} are the same to Eq.(10).

Based on Eq.(12), Eq. (13) can be rewritten as follows:

$$T_O + \frac{\left(\sum_{i=1}^{|RSO|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2) + R \cdot S \cdot R \cdot dpi^2\right) \cdot Bit \cdot CR}{E_j \cdot TR} \leq T_\theta^{Min} + \frac{T_\theta^{Max} - T_\theta^{Min}}{\delta} \cdot \left\lceil \frac{\delta}{UAD + 1} \right\rceil \tag{14}$$

Solving Eq. (14), it can be derived below:

$$R \cdot dpi \leq \sqrt{\frac{\left(T_\theta^{Min} + \frac{T_\theta^{Max} - T_\theta^{Min}}{\delta} \cdot \left\lceil \frac{\delta}{UAD + 1} \right\rceil - T_O\right) \cdot E_j \cdot TR}{Bit \cdot CR} - \sum_{i=1}^{|RSO|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2)}{R \cdot S}} \tag{15}$$

To obtain a relatively high resolution of the non-RSO part, its pixel resolution can be approximately represented in Eq. (16).

$$R \cdot dpi \approx \sqrt{\frac{\left(T_\theta^{Min} + \frac{T_\theta^{Max} - T_\theta^{Min}}{\delta} \cdot \left\lceil \frac{\delta}{UAD + 1} \right\rceil - T_O\right) \cdot E_j \cdot TR}{Bit \cdot CR} - \sum_{i=1}^{|RSO|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2)}{R \cdot S}} \tag{16}$$

Next, we study how to choose an optimal ID of each IB replica among its Δ replicas. For the non-RSO part of the image, as the lower and upper bound of the dpi (i.e., D_L, D_U), the ID number(i) of the non-RSO(R) replica, and the granularity (Δ) are met in Eq.(17).

$$\Delta(E_U) = \left\lceil \frac{D_U - D_L}{\sqrt{\frac{(\theta_T - T_0) \cdot E_U \cdot TR}{Bit \cdot CR \cdot \left(\sum_{i=1}^{|\Omega|} \sum_{j=1}^{\alpha_i} SIB_{ij} \cdot S + R \cdot S\right)^{-D_U}}}} \right\rceil \tag{17}$$

$$\Delta(E_L) = \left\lceil \frac{D_U - D_L}{\sqrt{\frac{(\theta_T - T_0) \cdot E_L \cdot TR}{Bit \cdot CR \cdot \left(\sum_{i=1}^{|\Omega|} \sum_{j=1}^{\alpha_i} SIB_{ij} \cdot S + R \cdot S\right)^{-D_L}}}} \right\rceil$$

where Δ is a granularity value.

As i is an integer, solving Eq.(17), we have:

$$i = \left\lceil \frac{(R \cdot dpi - D_L) \cdot \Delta}{D_U - D_L} \right\rceil \tag{18}$$

Combing Eq. (18) with Eq. (16), the replica ID for the non-RSO part can be derived as:

$$i = \left[\left(\sqrt{\frac{\left(T_{\theta}^{Min} + \frac{T_{\theta}^{Max} - T_{\theta}^{Min}}{\delta} \cdot \left\lceil \frac{\delta}{UAD + 1} \right\rceil - T_0 \right) \cdot E_j \cdot TR}{Bit \cdot CR} - \sum_{i=1}^{|RSO|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2)}{R \cdot S} - D_L} \right) \cdot \frac{\Delta}{D_U - D_L} \right] \quad (19)$$

Algorithm 2 summarizes a dynamic optimal RIB replica selection process in which two cases are considered, i.e., with *user tapping* (lines 7–10) and without *user tapping* (lines 2–5).

Algorithm 2. Dynamic Optimal RIB Replica Selection

Input: all of the RIB replicas, the transmission image I_S , network bandwidth E_j , and users' interactions

Output: the optimal ID for RIB replica

1. **while** (the transmission is not finished)
2. **if** No User Tapping **then**
3. **for** each RIB_i in I_S **do**
4. $RIB_i.dpi \approx \sqrt{\frac{T_{\theta} \cdot E_j \cdot TR}{Bit \cdot CR} - \sum_{i=1}^{|RSO|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2)}{R \cdot S}}$;
5. $i = \left[\left(\sqrt{\frac{T_{\theta} \cdot E_j \cdot TR}{Bit \cdot CR} - \sum_{i=1}^{|RSO|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2)}{R \cdot S} - D_L \right) \cdot \frac{\Delta}{D_U - D_L} \right]$;
6. **end for**
7. **else**
8. **for** each RIB_i in I_S **do** // at the slave node

$$9. \quad RIB_i.dpi \approx \sqrt{\frac{\left(T_{\theta}^{Min} + \frac{T_{\theta}^{Max} - T_{\theta}^{Min}}{\delta} \cdot \left\lceil \frac{\delta}{UAD + 1} \right\rceil - T_0 \right) \cdot E_j \cdot TR}{Bit \cdot CR} - \sum_{i=1}^{|RSO|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2)}{R \cdot S}}$$

$$10. \quad i = \left[\left(\sqrt{\frac{\left(T_{\theta}^{Min} + \frac{T_{\theta}^{Max} - T_{\theta}^{Min}}{\delta} \cdot \left\lceil \frac{\delta}{UAD + 1} \right\rceil - T_0 \right) \cdot E_j \cdot TR}{Bit \cdot CR} - \sum_{i=1}^{|RSO|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2)}{R \cdot S} - D_L \right) \cdot \frac{\Delta}{D_U - D_L} \right];$$

11. **end for**
12. **end if**
13. **return** i ;
14. **end while**

4.3 Optimal Δ

Since the value of Δ can affect the storage and maintenance costs for the replicas to some extent. To obtain an optimal Δ , in this subsection, we proceed to derive an optimal Δ .

Specifically, let us first denote the average resolution of the image I_S as $dpi \in [D_L, D_U]$, where D_L and D_U denote the lower and upper bound dpi of I_S , respectively. For description simplicity, we use image fragment (IF) to represent the $RSOs$ and R in the image. Since the IFs are composed of some IBs , so the pixel resolution of the IB replicas can be obtained by that of

the *IFs*. Therefore, the data sizes of the *IFs* and their corresponding pixel resolutions can be approximately met in Eq.(20):

$$\sum_{i=1}^{|RSO|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2) + R.S \cdot R \cdot dpi^2 = \left(\sum_{i=1}^{|RSO|} RSO_i \cdot S + R.S \right) \cdot dpi^2 \tag{20}$$

where *R.S* means the area of the non-RSO region.

Solving Eq.(20), the average *dpi* of the whole image can be derived in Eq. (21):

$$dpi = \sqrt{\frac{\sum_{i=1}^{|RSO|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2) + R.S \cdot R \cdot dpi^2}{\sum_{i=1}^{|RSO|} RSO_i \cdot S + R.S}} \tag{21}$$

In addition, the bandwidth of the *j*-th edge is defined as: $E_j \in [E_L, E_U]$, where E_L and E_U are the lower and upper bound of the bandwidth of the *j*-th edge, respectively. Note that, the above bandwidth is a theoretical value that is larger than the actual one. For the current network bandwidth E_j , we have $E_j \in \left[E_L + \frac{(i-1)(E_U-E_L)}{\Delta}, E_L + \frac{i(E_U-E_L)}{\Delta} \right]$. Since *i* is an integer, so $i = \left\lceil \frac{(E_j-E_L) \cdot \Delta}{E_U-E_L} + 1 \right\rceil$.

As illustrated in Fig. 6, based on an assumption that given a transmission time deadline T_θ , when the mobile network can't provide enough bandwidth, the image transmission processing may not be efficiently and successfully completed in T_θ . On the contrary, when the network bandwidth is sufficient enough, the average *dpi* of an image is proportional to the network bandwidth (E_j), the corresponding *dpi* of I_S under the current network bandwidth (E_j) can be derived as follows:

$$dpi = D_L + \frac{i \cdot (D_U - D_L)}{\Delta} \tag{22}$$

$$dpi = D_L + \left\lceil \frac{(E_j - E_L) \cdot \Delta}{E_U - E_L} + 1 \right\rceil \cdot \frac{D_U - D_L}{\Delta} \tag{23}$$

where $i \in [1, \Delta]$.

In Eq. (23), a whole image is regarded as an object to be processed. The pixel resolution of the whole image can be adjusted according to the variance of the network bandwidth. This method, however, may decrease the pixel resolution of the RSOs so much that the user cannot clearly view it. Therefore, in the preprocessing step, as shown in Fig. 5a, the RSOs in the

Fig. 6 Relationship between E_j and *dpi*

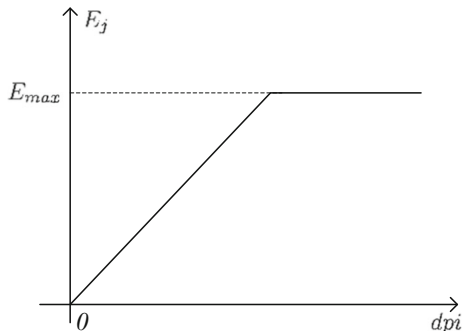


image are firstly identified by the red solid line rectangles automatically [11], i.e., RSO_1 . Figure 5b shows that the resolutions of the RSO in Fig. 5a and b are fixed, and the resolution of the rest part (R) in Fig. 5b, however, is decreased significantly.

Combing Eqs. (20) and (23), we have,

$$\sqrt{\frac{\sum_{i=1}^{|RSO|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2) + R \cdot S \cdot R \cdot dpi^2}{\sum_{i=1}^{|RSO|} RSO_i \cdot S + R \cdot S}} = D_L + \left[\frac{(E_j - E_L) \cdot \Delta}{E_U - E_L} + 1 \right] \cdot \frac{D_U - D_L}{\Delta} \tag{24}$$

Δ can be approximately derived by solving Eq.(24):

$$\Delta = \left[\frac{D_U - D_L}{\sqrt{\frac{\sum_{i=1}^{|RSO|} (RSO_i \cdot S \cdot RSO_i \cdot dpi^2) + R \cdot S \cdot R \cdot dpi^2}{\sum_{i=1}^{|RSO|} RSO_i \cdot S + R \cdot S}} - D_L - \left[\frac{(E_j - E_L) \cdot (D_U - D_L)}{E_U - E_L} \right]} \right] \tag{25}$$

To obtain an optimal Δ , Δ should be minimized such that the storage cost of IB replicas is minimal. So, let $RSO_i \cdot dpi = R \cdot dpi = dpi$, we have:

$$\Delta_{opt} = \left[\frac{D_U - D_L}{dpi - D_L - \left[\frac{(E_j - E_L) \cdot (D_U - D_L)}{E_U - E_L} \right]} \right] \tag{26}$$

From a theoretical perspective, the network bandwidth (i.e., E_j) varies in all bandwidth (i.e., $[E_L, E_U]$) ranging from $E_L = 10$ MB/S to $E_U = 100$ MB/S. In most real-life applications, however, E_j is relatively stable (i.e., varies in a small range $[E'_L, E'_U]$, where $E'_L \geq E_L$ and $E'_U \leq E_U$). Then,

$$\Delta_{opt} \in \left[\left[\frac{D_U - D_L}{dpi - D_L - \left[\frac{(E_L - E_L) \cdot (D_U - D_L)}{E_U - E_L} \right]} \right], \left[\frac{D_U - D_L}{dpi - D_L - \left[\frac{(E_U - E_L) \cdot (D_U - D_L)}{E_U - E_L} \right]} \right] \right] \tag{27}$$

Based on Eq.(27), the optimal Δ can be approximately represented as:

$$\Delta_{opt} \approx \left[\frac{D_U - D_L}{dpi - D_L - \left[\frac{(E_U - E_L) \cdot (D_U - D_L)}{E_U - E_L} \right]} \right] \tag{28}$$

5 Optimal IB replicas placement scheme

As data placement and storage is critically important for the data transmission, so in this section, to better facilitate the IIT processing, we propose an optimal IB replicas placement scheme based on the available network bandwidth.

For example, in Fig. 7, assume that there are three RSOs (e.g., RSO_1 , RSO_2 , and RSO_3) represented by the blue rectangles in an image. The image is first equally partitioned into 56 IBs in which the 9 IBs, 9 IBs, 6 IBs and 32 IBs belong to the RSO_1 , RSO_2 , RSO_3 , and R , respectively. The $SIBs$ are stored with the original resolution and the $RIBs$ are stored with different pixel resolutions based on Δ .

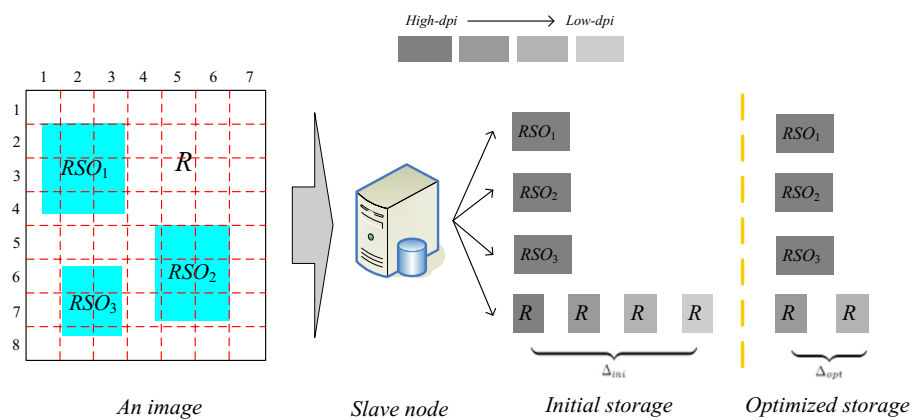
Based on Eq.(25), different network bandwidths (i.e., $E_j \in [E_L, E_U]$) correspond to different granularities. The initial granularity (Δ_{ini}) can be represented by Eq.(29):

$$\Delta_{ini} = \operatorname{argmax}\{\Delta(E_L), \Delta(E_U)\} \tag{29}$$

where $\Delta(E_U) = \left\lceil \frac{D_U = D_L}{\sqrt{\frac{(E_U - E_L) E_U E_X}{\sum_{i=1}^{|I|} \sum_{j=1}^{|S|} SIB_{ij} \cdot S + R \cdot S}}^{-D_U}} \right\rceil$, and $\Delta(E_L) = \left\lceil \frac{D_U = D_L}{\sqrt{\frac{(E_U - E_L) E_U E_X}{\sum_{i=1}^{|I|} \sum_{j=1}^{|S|} SIB_{ij} \cdot S + R \cdot S}}^{-D_L}} \right\rceil$.

Since the pixel resolutions of the $SIBs$ are fixed, next we focus on the study of the optimal granularity value for the RIB replicas. In Algorithm 1, the initial granularity value of the RIB replicas (Δ_{ini}) is designed for all bandwidth which ranges from $E_L = 10$ MB/S to $E_U = 100$ MB/S. The increase of Δ_{ini} , however, leads to a larger storage cost of the IB replicas. In most real-life applications, the network bandwidth (i.e., E_j) is relatively stable (i.e., varies in a small range) and the RSOs in the images are identified previously, which motivate us to investigate an optimal RIB replicas storage scheme based on the available network bandwidth. Therefore, to further reduce the storage cost of the RIB replicas, we propose a batch updating algorithm for the optimal RIB replicas storage.

First, suppose that the small range of the network bandwidth is denoted as $[E'_L, E'_U]$, combining with Eqs.(23) and (25), the optimal ID is derived below:



- $RSO_1 = 21, 22, 23, 31, 32, 33, 41, 42, 43$
- $RSO_2 = 54, 55, 56, 64, 65, 66, 74, 75, 76$
- $RSO_3 = 62, 63, 72, 73, 82, 83$
- $R = 11, 12, 13, 14, 15, 16, 17, 24, 25, 26, 27, 34, 35, 36, 37, 44, 45, 46, 47, 51, 52, 53, 57, 61, 67, 71, 77, 81, 84, 85, 86, 87$

Fig. 7 IB replicas placement at the slave node level

$$i_{opt} \approx \left\lceil \frac{RIB \cdot dpi - D_L}{\sqrt{\frac{(\theta_T - T_0) \cdot E_j \cdot TR}{Bit \cdot CR \cdot \left(\sum_{i=1}^{|\Omega|} \sum_{j=1}^{\alpha_i} SIB_{ij} \cdot S + R \cdot S \right)}^{-D_L} - \frac{(E_j - E_L) \cdot (D_U - D_L)}{E_U - E_L}}} \right\rceil \tag{30}$$

where $E'_L \geq E_L$ and $E'_U \leq E_U$.

Based on Eq.(30), the optimal replica IDs for the RIBs range from $[i_{opt}(E'_L), i_{opt}(E'_U)]$, if $i_{opt}(E'_L) \leq i_{opt}(E'_U)$; otherwise, $[i_{opt}(E'_U), i_{opt}(E'_L)]$. That means the replicas having granularity IDs not in the above ranges can be removed. Algorithm 3 summarizes a batch updating method for the optimal storage of the RIB replicas.

Algorithm 3. Batch updating of IB replicas()

Input: the initial IB replicas placement, n receiver users

Output: the optimal IB replicas placement

1. **if** $i_{opt}(E'_L) \leq i_{opt}(E'_U)$ **then**
2. the RIB replicas whose granularity IDs are not in $[i_{opt}(E'_L), i_{opt}(E'_U)]$ are removed;
3. **else**
4. the RIB replicas whose granularity IDs are not in $[i_{opt}(E'_U), i_{opt}(E'_L)]$ are removed;
5. **end if**

6 The IIT algorithm

With the support of the above enabling techniques, an image can be efficiently and interactively transferred in the MWN. Before introducing the IIT algorithm, we first provide a system overview of the transmission method.

6.1 System overview

Figure 8 shows the system overview of the IIT processing. Given an image I_S , the RSOs detection processing is first conducted by the approach of Girshick et al. [11]. Then, the image is physically and equally partitioned into several IBs that include two types: SIB and RIB. For the SIB replicas, they can be transmitted to the receiver with their original resolutions at top priorities. For the RIB replicas, their pixel resolutions and transmission priority are lower than

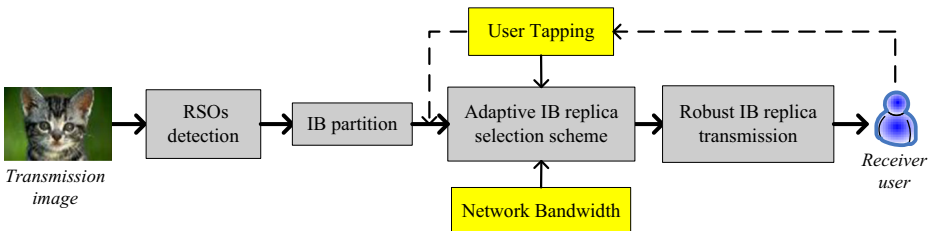


Fig. 8 System overview of the IIT processing

that of the *SIBs*, so that the key part of the image can be transmitted and displayed in priority. Once the preprocessing step is completed, the next step is to perform the *IIT* processing. Specifically, when an image I_S is prepared to transmit, the information such as current network bandwidth and the receiver user interaction behaviors need to be collected and analyzed uniformly to derive the optimal pixel resolutions of the *RIB* replicas. Finally, based on their transmission priorities, the candidate IBs are transmitted to the receiver node (N_R) where the IBs are reconstructed and displayed.

6.2 The algorithm

Algorithm 4 summarizes the detailed steps of our proposed *IIT* algorithm. First of all, a user U_R sends an image transmission request (i.e., the ID number of I_S) to the sender node level N_S (line 1), then the RSOs in the image are first identified (line 2). After that, the interactive transmission processing starts (line 3). In line 4, *SIB* replicas are first sent to the sender node with their original resolution. After that, for the *RIB* replicas transmission, we perform the dynamic optimal *RIB* replica selection processing to obtain the optimal ID of *RIB* replica (line 5), then sent to the sender node (line 6).

Algorithm 4. *IIT*(I_S, U_R)

Input: I_S : a transmission image and some parameters, a user U_R

Output: the image transferred

1. an image transmission request is submitted from the receiver node N_R by the user U_R ;
 2. **if** the RSOs have not been identified in I_S **then** identify the RSOs in this image;
 3. **while** (the image transmission is not finished)
 4. send the *SIB* replicas to the sender node with their original resolutions;
 5. perform the dynamic optimal *RIB* replica selection processing; // see algorithm 2
 6. send the *RIB* replica to the sender node based on the TP in a descending order;
 7. **end while**
-

7 Experimental evaluation

To demonstrate the efficiency and reliability of our proposed *IIT* method, we conduct simulation experiments to demonstrate the transmission performance.

7.1 Experiment setup

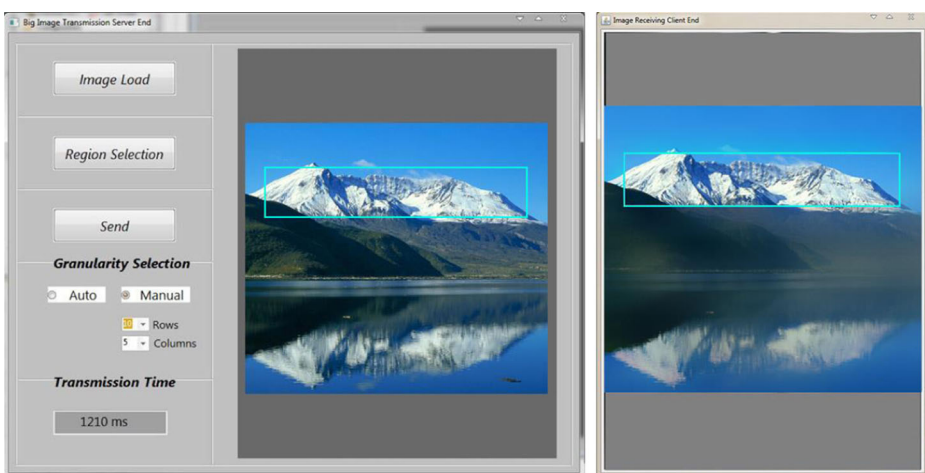
The prototype image receiver client is implemented on Android platform [25] in the Java language. The user interactions (i.e., tapping the screen) are randomly simulated by the users. Each node has a 2.7GHz Xeon processor, 2.0GB memory and 1 TB hard disk. The nodes in the local area network are connected via 1Gbps network links. The number of nodes in our

system varies from 10 to 100. All experiments are performed in a 4G cellular network in which the average and maximum data communication rates are 10 bps and 30Mbps, respectively. In the slave nodes, the IB replicas with different pixel resolutions are stored in a file system and some information is recorded in a MySQL [20] database.

- **Datasets.** To objectively and extensively evaluate the *IIT* method, we adopt two image datasets that are obtained from two ways: 1) **Real dataset:** 100,000 images are downloaded from Internet in which the image data size ranges from 0.2 to 1 MB; 2) **Synthetic dataset:** To evaluate the effect of data size on the image transmission performances, we have synthesized five groups of image data in which the data size of each image are 1 MB, 5 MB, 10 MB, 50 MB and 100 MB, respectively.
- **Competitors.** There are two competitors in our experiments. The first one is a baseline - *traditional transmission method* (i.e., transmitting the whole image without partitioning); the second one is the CBMR scheme [30].

7.2 A prototype transmission system

We have implemented a prototype transmission system for images as illustrated in Fig. 9. Figure 9a shows an example of the backend interface of the offline image processing. One RSO in this figure has been identified by a blue rectangle. Figure 9b demonstrates the receiver client interface in which the IBs in the two IFs (e.g., one *RSO* represented by the blue-line rectangle and the rest part *R*) have been reconstructed and displayed. Moreover, the pixel resolution of the RSO is higher than that of the rest part.



(a) A backend interface

(b) A client end interface

Fig. 9 An interactive image transmission system

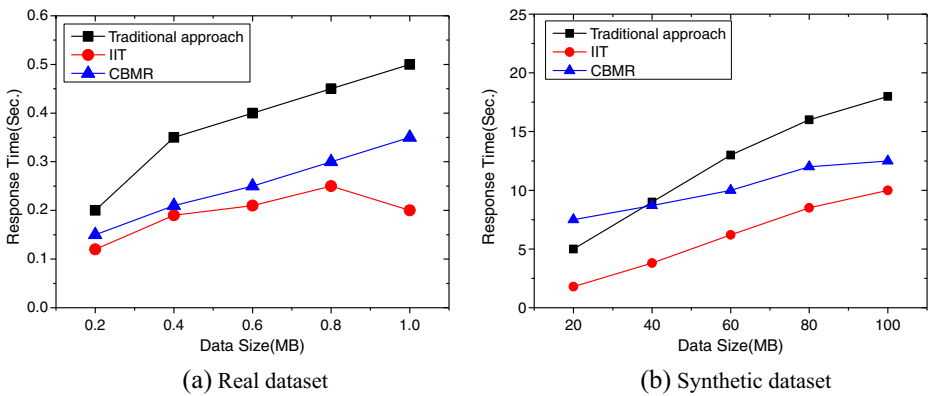


Fig. 10 Effect of image size

7.3 Effect of image size

In the first experiment, we study the effect of the image size on the performance of the *IIT* processing by using two kinds of image data as mentioned above. Method 1 uses traditional transmission method (i.e., transmitting the whole image without partitioning), method 2 adopts the CBMR scheme [30], while method 3 adopts the *IIT*. In Fig. 10a and b, when the bandwidth (e.g., 100 MB/Sec) is relatively stable, the total transmission time using the *IIT* is better than that of other two ones. Meanwhile, with the increase of the image data size, the performance gap of the two approaches increases with the image size. This is because compared with the *IIT*, the corresponding data size of the images to be transmitted with the traditional approach is increasing so rapidly that the images cannot be sent to the destination nodes quickly. Our hybrid pixel resolution approach can effectively reduce the transmitted image data, especially for a large image.

7.4 Effect of network bandwidth

Next, we investigate the effect of the network bandwidth on the performance of the *IIT* processing by using the two kinds of images. The three methods we compared with are the

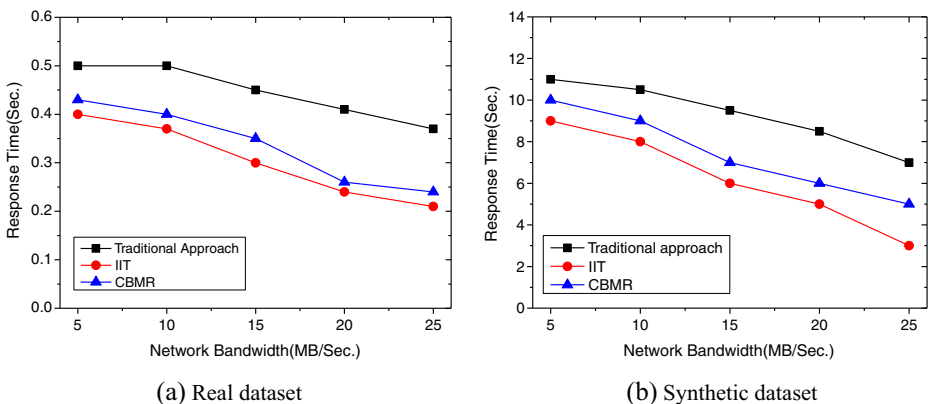


Fig. 11 Effect of network bandwidth

Table 2 The description of the opinion scores

<i>OS</i>	1	2	3	4	5
<i>Quality</i>	Bad	Poor	Fair	Good	Excellent

same to the ones in Section 7.3. Figure 11 shows when the image data sizes (e.g., 0.5 MB for real data and 50 MB for the synthetic data) are fixed, the total response time using our proposed *IIT* method is superior to that of other two methods. Meanwhile, with the condition that the bandwidth is increasing, the response time decreases gradually and the performance gap becomes larger especially for the large synthetic image data. This is because in the *IIT* method, the original image data size has been significantly reduced based on the three factors, i.e., network bandwidth, the image contents, and user interaction.

7.5 Effect of Δ

In this experiment, we proceed to test the effect of Δ on the transmission cost, storage cost, and the mean opinion score(*MOS*) defined in Eq.(31), respectively.

$$MOS = \frac{\sum_{i \in [1, |U_i|]} U_i \cdot OS}{|U_i|} \tag{31}$$

where $U_i \cdot OS$ means the opinion score(*OS*) for the user U_i (ref. Table 2) and $|U_i|$ is the total number of the users.

According to Eq.(31), 100 volunteers are randomly selected and involved to decide whether the image replica is clear to view by answering ‘*excellent*’, ‘*good*’, ‘*fair*’, ‘*poor*’, or ‘*bad*’. The synthetic image data set is used to perform the experiment.

In Fig. 12a, with the increase of Δ , the transmission cost is gradually decreasingly. This is because when Δ is small, it is hard to find the suitable IB replicas since the gap between the data sizes of the current IB replicas and the optimal ones becomes larger, leading to a higher transmission cost. Similarly, as illustrated in Fig. 12b, the storage cost increases rapidly as Δ increases since the total number of the IB replicas is increasing when Δ becomes larger, which leads to a larger storage cost of the IB replicas. Finally, we study the effect of Δ on *MOS*. In Fig. 12c, when Δ is larger than 20, the *MOS* is increased no more. This is because when Δ is too small, the image quality of the image reconstruction becomes relatively low which results in a poor user experience. So to obtain a tradeoff between transmission cost, storage cost and image quality, an optimal number of Δ is found to be 20.

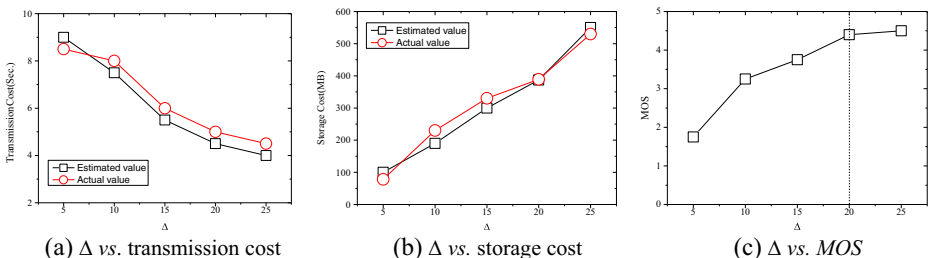
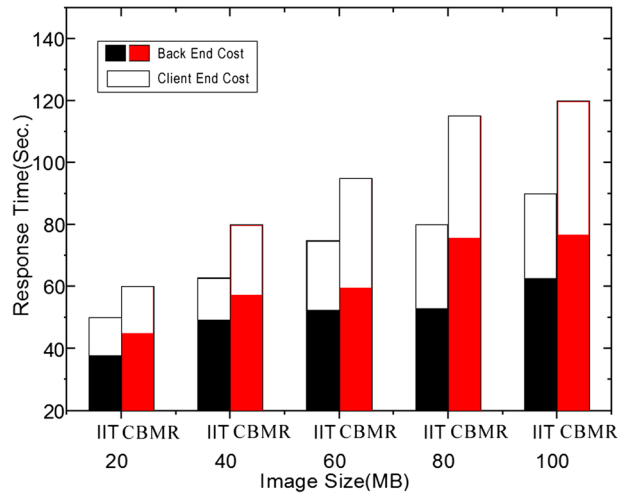


Fig. 12 Effect of Δ

Fig. 13 Back end vs. client end

7.6 Back end vs. client end

In this experiment, we compare the overheads of the backend and the client end of our transmission method and the CBMR [30] in which 50 transmission requests are randomly generated and conducted to obtain the average consuming times for the above two ends respectively. We adopt the synthetic data as the experimental one. Figure 13 illustrates that when the image data size is from 20 MB to 100 MB, the transmission time of the *IIT* is gradually increasing and smaller than that of the CBMR. Additionally, the overheads of the backend and client end for the CBMR method are slightly larger than the *IIT*. This is because for the CBMR, its computational costs for the image compression in the back end are larger than the *IIT*. The decoding costs for the CBMR in the client end are also larger than the *IIT*.

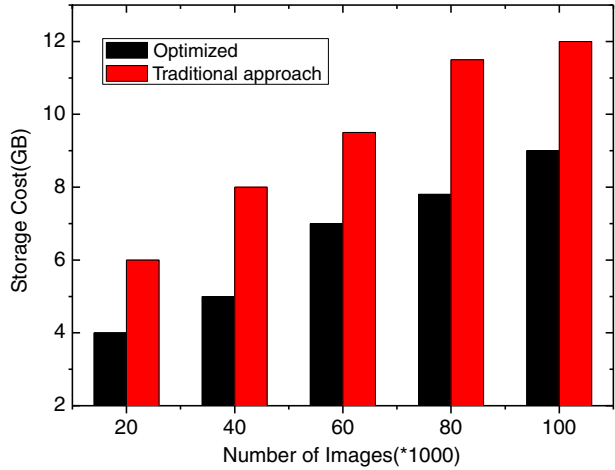
7.7 Effect of optimized storage scheme

This experiment evaluates the effect of storage scheme by adopting the real data as the experimental one. Figure 14 shows that when the number of images increases, the storage cost of the traditional approach is much larger than that of the optimized one. This is because for the optimized storage scheme, the data size of the image replicas is significantly reduced when the network bandwidth varies in a small range.

7.8 Effect of the TPA

In the last experiment, we first test the efficiency of the two transmission schemes: 1) *Our proposed TPA method* and 2) *the CBMR approach* [30] by using the synthetic dataset. In Fig. 15a, when the image data size is from 20 MB to 100 MB, the transmission time of TPA is gradually increasing and better than that of the CBMR. This is because the data compression rate of the TPA method is higher than the CBMR one. The data size to be transferred by the CBMR is larger than that of the TPA.

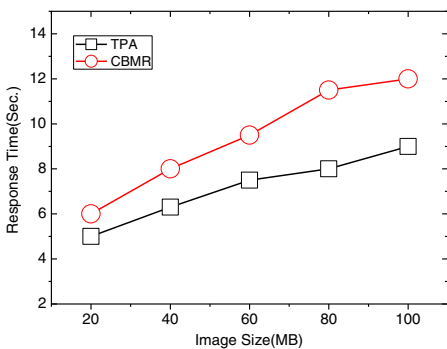
Fig. 14 Effect of storage scheme



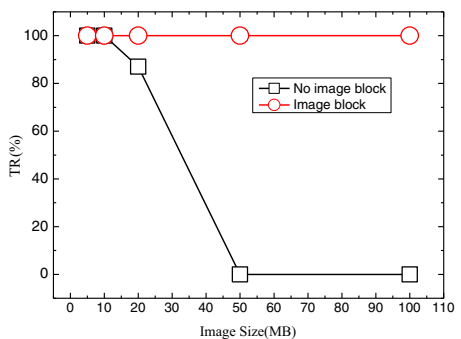
To evaluate the effect of image size on the transmission robustness, we then used the synthetic image dataset in which the images have been divided into five groups in terms of the data size such as 5 MB, 10 MB, 20 MB, 50 MB and 100 MB. The transmission reliability (TR) can be defined in Eq.(32).

$$TR = \frac{\text{Number of successful data transmissions}}{\text{Total number of data transmissions}} \tag{32}$$

As shown in Fig. 15b, with the increase of data size, the successful data transmission ratio is 100 % by using our image blocking technique. For the data transmission without adopting the TPA method, when the transmission data size is less than 10 MB, the successful transmission ratio is 100 %, however, if the data size is 20 MB, the average TR is decreased to 87 %. And if the data size is larger than 50 MB, the average TR is zero since it is hard to transmit such a large image successfully in a MWN . Based on the experimental result, to guarantee a high successful data transmission ratio, it is possible to transfer a large image only through the image blocking method in a limited network bandwidth.



(a) Image size vs. transmission efficiency



(b) Image size vs. transmission robustness

Fig. 15 Effect of TPA

8 Conclusions and future work

In this paper, we have presented an interactive large image transmission scheme (*IIT*) in the resource-constraint *MWN*, which allows the users to actively affect the transmission processing through tapping the receiver's screen. Two enabling techniques, i.e., *multi-resolution-based IB replica selection scheme* and *optimal IB replicas placement scheme* are proposed to optimize the image transmission. Our experiments demonstrate that our proposed *IIT* method is more suitable for the large image transmission in minimizing the network communication cost as well as maximizing the rate of successful image transmission and user satisfaction.

In the future, to further enhance the transmission efficiency of the multiple images in a transmission-intensive environment, we plan to study a batch image transmission scheme based on the *IIT* approach. For real-life applications, we are planning to apply our application into of the traffic anomaly detection [14] and energy-efficient multicast routing [13, 15].

Acknowledgments The authors would like to thank the editors and anonymous reviewers for their helpful comments. This work is partially supported by the Program of National Natural Science Foundation of China under grant No. 61272188, 61379075, 61540064, 71571162; the project in National Science & Technology Pillar Program of the Ministry of Science and Technology under grant No. 2014BAK14B01; the Program of Natural Science Foundation of Zhejiang Province under grant No. LY13F020008; the Ministry of Education of Humanities and Social Sciences Project under grant No. 14YJCZH235; the “Qianjiang Talent” Project of Zhejiang Province under grant No. QJD1402017.

References

1. Allcocka B, Bestera J, Bresnahan J etc (2002) Data management and transfer in high-performance computational grid environments. *Parallel Comput* 28(5), 749–771
2. Arslan SS, Cosman PC, Milstein LB (2012) Generalized unequal error protection LT Codes for progressive data transmission. *IEEE Trans Image Process* 21(8):3586–3597
3. Boluk PS, Baydere S, Emre Harmanci A (2011) Robust Image Transmission Over Wireless Sensor Networks. *J Mob Netw Appl* 16(2):149–170
4. Chang CC and Wu MN (2003) A color image progressive transmission method by common bit map block truncation coding approach. In *Int'l Conf Commun Technol* (2), 1774–1778
5. Chang CC, Shine FC, Chen TS (1999) A new scheme of progressive image transmission based on bit-plane method. *Asia-Pacific Conf Commun Fourth Optoelectron Commun Conf* 2:892–895
6. Chang CC, Shih TK, and Lin IC (2002) An efficient progressive image transmission method based on guessing by neighbors. *Vis Comput* (18), 341–353
7. Chang RC, Shih TK, Hsu HH (2008) A Strategic Decomposition for Adaptive Image Transmission. *J Inf Sci Eng* 24(3):691–707
8. Charles JT, Larry LP (1992) Image transfer: an end-to-end design. *ACM SIGCOMM Int'l Conference on Data Communication*:258–268
9. Gao DH, Liu DH, Feng YQ et al (2010) A Robust Image Transmission Scheme for Wireless Channels Based on Compressive Sensing. *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence. Lect Notes Comput Sci* 6216:334–341
10. Gelogo YE, Kim T-h (2013) Compressed Images Transmission Issues and Solutions. *Int J Comput Graph* 5(1):1–8
11. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*
12. Hu Y, Xie X, Chen Z, Ma W-Y (2004) Attention Model Based Progressive Image Transmission. *Proc of IEEE Int Conf Multimedia Expo* 2:1079–1082
13. Jiang D, Xu Z, Li W, Chen Z (2015) Network coding-based energy-efficient multicast routing algorithm for multi-hop wireless networks. *J Syst Softw* 104(2015):152–165
14. Jiang D, Yuan Z, Zhang P, Miao L, Zhu T (2016a) A traffic anomaly detection approach in communication networks for applications of multimedia medical devices. *Multimedia Tools and Applications*, online available

15. Jiang D, Shi L, Zhang P, Ge X (2016b) QoS constraints-based energy-efficient model in cloud computing networks for multimedia clinical issues. *Multimedia Tools and Applications*, online available
16. John MD, Georey MD, Song XY (1995) Fast lossy internet image transmission. In *ACM Int'l Conf Multimedia*
17. Kim JH and Song WJ (1996) Pyramid-structured progressive image transmission using quantization error delivery in transform domains, *IEE Vision, Image and Signal Processing*. (143), 132–136
18. Lin T, Hao P (2005) Compound image compression for real-time computer screen image transmission. *IEEE Trans on Image Process* 14(8):993–1005
19. Maani R, Camorlinga S, Arnason N (2012) A parallel method to improve image transmission. *J Digit Imaging* 25(1):101–109
20. MySQL (2010). <http://www.mysql.com/>
21. Raman S, Balakrishnan H, Srinivasan M (2000) An image transport protocol for the Internet. *Int'l Conf Network Protocol*:209–219
22. Ruiz VG, Fernández JJ, and García I (2001) Image compression for progressive transmission. In the Nineteenth IASTED Int'l Conference on Applied Informatics: Advances in Computer Applications. 519–524. Innsbruck, Austria
23. Aziz SM Pham DM (2013) Energy Efficient Image Transmission in Wireless Multimedia Sensor Networks. *IEEE Commun Lett* 17(6):1084–1087
24. Sun Y, Xiong Z-X (2006) Progressive Image Transmission over Space-Time Coded OFDM-Based MIMO Systems with Adaptive Modulation. *IEEE Trans on Mobile Computing* 5(8):1016–1028
25. The Android platform (2010), www.google.com/android
26. Tzou KH (1987) Progressive image transmission: a review and comparison of techniques. *Opt Eng* 26:581–589
27. Victor S, Abugharbieh R, Nasiopoulos P (2010) 3-D scalable image compression with optimized volume of interest coding. *IEEE Trans on Medical Imaging* 10(29):1808–1820
28. Wu H, Abouzeid AA (2004) Power Aware Image Transmission in Energy Constrained Wireless Networks. *Proc of Ninth International Symposium on Computers and Communications* 1:202–207
29. Xua H, Hua K, Wang H (2015) Adaptive FEC coding and cooperative relayed wireless image transmission. *Digital Communications and Networks* Vol. 1, Issue 3, August pp. 213–221
30. Zhuang Y, Jiang N, Wu Z-A, Li Q etc (2014) Efficient and Robust Large Image Retrieval in Mobile Cloud Computing Environment. *Information Science*



Dr. Yi Zhuang is a recipient of the CCF Doctoral Dissertation Award conferred by Chinese Computer Federation in 2008 and IBM Ph.D. Fellowship 2007–2008. He is currently an associate professor at the College of Computer & Information Engineering in Zhejiang Gongshang University where he joined as faculty member since May 2008. He obtained his Ph.D. degree in computer science from Zhejiang University in March 2008. From January 2008 to March 2008, supported by IBM Ph.D. Fellowship, Dr. Zhuang has spent three months to participate in the study of an optimal hybrid storage model based on DB2 as a research intern in IBM China Research Lab. His research interests mainly focus on multimedia database, medical image retrieval and cloud computing. He has published 40+ papers in the leading journals and conferences, i.e., ACM TOIT, ACM TALIP, Information Sciences, WWW Journal, Multimedia Tools and Applications, Knowledge and Information Systems and Expert System with Applications, etc.



Nan Jiang received the bachelor degree of medical science and the master degree of medical science both from the Zhejiang University in 2004 and 2007, respectively. Her research interests mainly focus on medical image processing. She has published 10+ papers in international conferences and journals.



Hua Hu is a professor at School of Computer Science, Hangzhou Dianzi University where he joined as a faculty member since Dec. 2008. Before that, he taught at the Zhejiang Gongshang University. He received the bachelor degree, the master degree and the Ph.D degree from Zhejiang University in 1990, 1992 and 1998, respectively. His research interests are database systems and work flow. He has published 50+ international conference and journal papers.



Dr. Dickson K.W. Chiu received the B.Sc. (Hons.) degree in Computer Studies from the University of Hong Kong (HKU) in 1987. He received the M.Sc. (1994) and the Ph.D. (2000) degrees in Computer Science from the Hong Kong University of Science and Technology (HKUST). He started his own computer company while studying part-time. He has also taught at several universities in Hong Kong and served as a Research Fellow at the City University of Hong Kong. He is currently teaching at the HKU. His research interest is in e-business with a cross-disciplinary approach and the results have been published in over 160 papers in international journals and conference proceedings, including many practical master and undergraduate project results. He received a best paper award in the 37th Hawaii International Conference on System Sciences in 2004. He is the founding Editor-in-chief of the International Journal on Systems and Service-Oriented Engineering, International Journal on Organizational and Collective Intelligence, EAI endorsed Transaction on e-Business, and serves in the editorial boards of several international journals. He co-founded several international workshops and co-edited several journal special issues. He also served as a program committee member for over 100 international conferences and workshops. Dr. Chiu is a Senior Member of the ACM, a Senior Member of the IEEE, and a Life Member of the Hong Kong Computer Society.



Qing Li is a Professor at the Department of Computer Science, City University of Hong Kong where he joined as a faculty member since Sept. 1998. Prior to that, he has taught at the Hong Kong Polytechnic University, the Hong Kong University of Science and Technology and the Australian National University (Canberra, Australia). Prof. Li has served / is serving as a consultant to Microsoft Research Asia (Beijing, China), Motorola Global Computing and Telecommunications Division (Tianjin Regional Operations Center), and the Division of Information Technology, Commonwealth Scientific and Industrial Research Organization (CSIRO) in Australia. He is a Guest Professor of the University of Science and Technology of China (USTC) and Zhong Shan (Sun Yat-Sen) University, a Visiting Professor at the Institute of Computing Technology (Knowledge Grid), Chinese Academy of Science (Beijing, China), and was an Adjunct Professor of the Huazhong University of Science and Technology and the Hunan University (Changsha, China) where he got his BEng. degree from the Department of Computer Science in 1982. He is also a Guest Professor (Software Technology) of the Zhejiang University (Hangzhou, China). Prof. Li has been actively involved in the research community by serving as an associate

editor and reviewer for technical journals, and as an organizer/co-organizer of numerous international conferences. Some recent conferences in which he is playing or has played major roles include WAIM2010, DASFAA2010, APWeb-WAIM'09, ER'08, WISE'07, ICWL'06, HSI'05, WAIM'04, IDEAS'03, VLDB'02, PAKDD'01, IFIP 2.6 Working Conference on Database Semantics (DS-9), IDS'00, and WISE'00. In addition, he served as a programme committee member for over fifty international conferences (including VLDB, ICDE, WWW, ER, CIKM, CAiSE, DASFAA, CoopIS, and FODO). He is currently a Fellow of IET (formerly IEE), a Senior Member of IEEE, a member of ACM-SIGMOD and IEEE Technical Committee on Data Engineering. He is the chairperson of the Hong Kong Web Society, and also served/is serving as an executive committee (EXCO) member of IEEE-Hong Kong Computer Chapter and ACM Hong Kong Chapter. In addition, he serves as a councilor of the Database Society of Chinese Computer Federation, a councilor of the Computer Animation and Digital Entertainment Chapter of Chinese Computer Imaging and Graphics Society, and is a Steering Committee member of ER, DASFAA, ICWL, UMEDIA, and the international WISE Society.